DOCUMENT RESUME

ED 294 906 TM 011 578

AUTHOR Murphy, Edward J.; And Others
TITLE Bias Review Methods for a Teacher Certification Testing Program.
PUB DATE Apr 88
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Teacher Certification; Teacher Evaluation; *Test Bias; Test Construction; *Testing Programs
IDENTIFIERS *Illinois Certification Testing System

ABSTRACT

Bias prevention activities undertaken as part of the development process for a major teacher certification testing program in Illinois are described. Because bias concerns were a major issue in planning and implementing the program, three distinct committees were formed to help prevent bias: (1) a bias review committee; (2) a content advisory committee; and (3) a technical review committee. Both statistical and judgmental procedures were used to detect bias, but the small pools of examinees in many fields limited statistical applications. Judgments of educators were the only means of detecting bias in many cases. Additional statistical methods were planned as the program moved into full implementation. An appendix contains the test field codes for the certification testing system. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED294906

BIAS REVIEW METHODS FOR A TEACHER CERTIFICATION TESTING PROGRAM

Edward J. Murphy
National Evaluation Systems, Inc.

Paula M. Nassif
National Evaluation Systems, Inc.

Susan K. Bentz
Illinois State Board of Education

Paper Presented at the Annual Meeting of the
American Educational Research Association
New Orleans, Louisiana
April 1988

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

EDWARD J. MURPHY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM011578

INTRODUCTION

Preventing bias in testing programs is important. In teacher certification programs, in which the stakes for examinees are high, bias prevention is of special concern. This paper describes bias prevention activities undertaken as part of the development process for a major teacher certification testing program in Illinois.

The approach taken in this paper is qualitative and descriptive. We will describe activities and procedures that were undertaken as prudent measures for meeting an established goal: preventing bias in the development of the Illinois program. In attempting to meet that goal, those who are involved in the development process have gone through a number of useful and instructive experiences. We feel that these experiences are of sufficient interest to present them for potential consideration and use by others engaged in actual high-stakes test development.

This paper is organized as follows. After a brief discussion on statistical and judgmental approaches to bias prevention in general, the paper describes the bias prevention activities that have been an integral part of the development process for the Illinois Certification Testing System.

This presentation is followed by a discussion of some conclusions drawn by the authors about the particular approaches used in the Illinois program, their possible significance, and their implications for similar programs elsewhere.

The Illinois program is at this point still under development; the first administration of the program is scheduled for July 1988. Therefore, our description of the bias prevention steps that have been undertaken for the Illinois program should be considered partial.

STATISTICAL AND JUDGMENTAL APPROACHES TO BIAS PREVENTION

Test program developers and administrators are guided in their efforts to ensure fairness in testing by the APA Standards for Educational and Psychological Testing (1985). Steps to eliminate bias are a fundamental part of a test development process. Bias prevention is grounded in the principles of sound test development and should be an integral part of test development activities.

Over the years, a number of approaches that deal with the issue of bias prevention have been developed and reviewed. Many books and handbooks have been produced that address various aspects of bias in test development, including, for example, Handbook of Methods for Detecting Item Bias

(Berk, 1982) and Bias Issues in Teacher Certification Testing (National Evaluation Systems, 1987). Some treatments of this issue deal with statistical techniques, others with developmental approaches, including judgmental techniques.

There are numerous ways to deal with bias as part of the development and/or administration of a testing program. In some programs, bias is reviewed at each step of the test development process and/or at specified points during program administration. Under this approach, bias reviews are considered central to test development activities and are performed on a continuing basis by those involved directly with the development process. Tittle (1982) recommends including minority representatives in review activities that are held throughout the test development process, from test planning through item writing, field tryout, item selection, and the development of norms and scales. In this approach, bias prevention is not accomplished by a "one-time review"; it is a consistent effort.

In other programs, bias review involves one meeting of persons external to the development process or a review (typically of test items) by a bias panel. Item statistics may or may not be available to reviewers depending on the point in the process at which the bias review takes place.

No matter what approach is taken, it is generally held that professional judgment has an important role. Because statistical results can be difficult to interpret, statistical techniques are rarely employed without

submitting the data to analysis by a group of experts or program administrators before final determinations are made.

Statistical models. Statistical techniques can flag differential performance on test items. Such flags serve a useful and important function: they help organize and highlight large bodies of educational materials (e.g., objectives, items) that need further review by content experts and/or bias reviewers. Although statistics alone, even those indicating adverse impact, do not of themselves determine bias, statistical differences do point to areas worthy of further consideration.

There are many statistical approaches for identifying potential item bias. Three approaches that have been used are item response theory, chi-square, and Angoff's delta-plot, or transformed item difficulty method (Shepard, Camilli & Williams, 1985; Ironson, 1982; Angoff, 1982). Comparisons of these methods show all to be useful but not absolute in detecting item bias. It is often the case that different statistical procedures flag different items, and results from statistical techniques are different from those of judgmental approaches. The statistical procedures that have been devised to date are relative measures, not absolute ones (Burrill, 1982).

Many statistical approaches require large numbers of examinees and, in particular, substantial numbers of minority examinees if statistical comparisons are to be made. Approaches that rely on item response theory generally require relatively large numbers; for programs with smaller

numbers of examinees, particularly smaller numbers of minority examinees, item response approaches are therefore ineffective. The chi-square method, while useful, can be difficult and time-consuming to implement and interpret. The Angoff transformed item difficulty, modified using residualized deltas, is often used (Shepard, et al., 1985). Recently, the Mantel-Haenszel approach (Holland & Thayer, 1986) has been favorably received. It too needs a substantial number of examinees for reliable comparisons.

Judgmental approaches. Shepard (1982) suggests a three-step approach to bias detection, including a logical analysis relating items to an implied trait, a statistical technique to identify aberrant items, and the use of judgment to identify patterns in the statistically biased items, which can be used to refine further the definition of the trait.

Tittle (1982) considers various types of bias in the context of ongoing bias detection in a program. The Illinois approach is consistent with Tittle's recommendations. It is specifically designed to identify and address potential bias at many points in the development process. This approach uses the expert judgments of many groups, supplemented where possible by statistical information.

Today, large programs often opt for a combination of statistical techniques and expert judgment. Many will combine bias checks throughout the process and independent reviews at specified points. The procedure employed in Illinois embodies this philosophy. This paper will describe in some detail the Illinois approach to bias prevention in the teacher certification test development process.

THE ILLINOIS CERTIFICATION TESTING SYSTEM

Background. The Illinois State Board of Education (ISBE) received a legislative mandate under the Education Reform Act of 1985 to effect several initiatives designed to improve professional standards for Illinois educational personnel. One of the initiatives is the Illinois Certification Testing System (ICTS).

The Reform Act stipulates that after July 1, 1988, individuals seeking early childhood, elementary, special, high school, school service personnel, or administrative certificates must pass a test of basic skills (comprising grammar, reading, mathematics, and writing subareas) and a test of subject-matter knowledge. The individuals included are new graduates from teacher preparation programs, educators moving to Illinois from other states, and Illinois educators applying for additional certificates.

The testing program is one component of a multifaceted qualifying process through which candidates for certificates must pass; candidates must also demonstrate competence in the usual ways at their institutions of higher education through the number of credit hours they have achieved in their desired fields within an approved program and through internships and job-experience requirements.

In spring 1986, the ISBE contracted with National Evaluation Systems, Inc. (NES) of Amherst, Massachusetts, to design and develop the 54 separate

tests that compose the ICTS program: one Basic Skills test and 53 subject-matter tests. (A list of the subject-matter test areas is appended.) This program is to be completed in time for a first administration in accordance with the legislatively mandated target of July 1, 1988.

Mandating bias prevention. Bias prevention has been a goal of the ICTS since its conception. The enabling legislation for the program, the Education Reform Act of 1985, requires that the "tests shall be designed to be racially neutral to assure that no person in taking the tests is thereby discriminated against on the basis of race, color, national origin or other factors unrelated to the person's ability to perform as a certificated employee." With such a mandate, the ISBE made bias concerns an important issue in the RFP for the program, during the selection of a contractor, as the design of the program became firm, and throughout the implementation of the program.

Illinois State Board of Education review committees. Early in the program, the ISBE arranged for the active involvement in the program of state-formed committees of three distinct types. Each committee would play an active role in efforts to prevent bias in the testing program.

1. A Bias Review Committee, consisting of Illinois educators with experience in assessing potential bias in educational materials, was formed especially for the ICTS. This committee was given responsibility for reviewing test development activities with the staff of the ISBE.

2. Content Advisory Committees were formed to review and edit draft written materials (e.g., objectives, items), advise the ISBE on content-related issues that might affect test design, and recommend the specific content of each test in the program. In all, 51 Content Advisory Committees were formed for the ICTS: one for the Basic Skills test and one for each of the distinct certificate fields (one committee served for the two levels--kindergarten through grade 12 and grades 6 through 12--of the Art, Music, and Physical Education fields). The Content Advisory Committee members were educators with content knowledge in the certificate field; they were drawn primarily from two populations: practitioners (i.e., elementary and high school educators) and college educators (i.e., faculty members involved in the preparation of educators). In addition, an Equity Resource Person, a member of the ISBE staff whose agency function pertains to equity and bias issues, was assigned as a member of each Content Advisory Committee.

The Content Advisory Committees not only reviewed and approved materials for content-related issues, they also were oriented in bias prevention so that as they gave detailed consideration to the draft materials they could review them for potential bias as well. The agency Equity Resource Persons played an important role in these reviews, but every member of the committee was asked to pay attention to the issue of fairness.

In assembling the 51 Content Advisory Committees, the ISBE considered it important to reflect the racial, ethnic, gender, and regional diversity of Illinois. For example, racial and ethnic minorities composed over 18 percent of the overall committee membership, and 24 percent of the committees were chaired by minorities.

3. A Technical Review Committee was formed to advise the ISBE on technical issues relating to the development of the ICTS and another testing program of the ISBE. The role of the Technical Review Committee was to review for their technical appropriateness proposed design features of the programs and make recommendations about design issues for consideration by the ISBE. Among the issues to be considered by the Technical Review Committee were the effects of program design elements on bias prevention and methods for detecting potential bias in the implementation of the program.

Thus, to prevent bias, the ISBE designed a structured, sequential approach whereby one committee could review, profit by, and extend the findings of a committee that had performed its review previously.

BIAS PREVENTION IN THE TEST DEVELOPMENT PROCESS

The ISBE worked with its contractor, NES, to design a process that reinforced the state's determination to prevent bias in the ICTS. Through each step in the multistep process that the program followed, bias remained a primary concern. The major steps in the test development process are described below.

1. Content definition. The fundamental method for preventing bias in a certification examination is to base that examination on important job-related content. To ensure that the content of each of the tests in the program would reflect actual job requirements and practice, the frameworks (i.e., objective sets structured into outline format) of the individual certificate test fields were drafted with reference to Illinois resources that related to the educator's job in each field. The resources that were used included the program descriptions from the teacher education programs that prepare candidates for each certificate, textbooks in use in Illinois classrooms and teacher preparation institutions, relevant state statutes and regulations, classroom curriculum materials, Illinois Student Assessment Program materials, professional literature, and similar documents. In addition, the ISBE curriculum specialist for each certificate field was consulted during the development of the frameworks and objectives.

These steps were taken to define domains of content for the tests that would reflect the fields as they are practiced in Illinois.

2. Orientation and training. Those who were to participate in the drafting of the content frameworks were oriented to the ICTS and given training in the development of objectives for certification tests. A significant part of the orientation and training concerned issues of bias. A manual on the topic, Bias Issues in Test Development, was updated for the Illinois program by National Evaluation Systems with the help of experts in the field of bias detection and prevention. This manual was issued to the objective developers, who were instructed to use the contents as they worked on objectives.

Early in the program, the manual was further modified and tailored for the Illinois program by members of the ISBE staff and the Bias Review Committee, who provided comments and suggestions for additions and improvements to it.

One important feature of the manual was its definition of bias through the identification of four categories. These are listed and briefly described here because they were salient in all bias review activities conducted for the ICTS and influenced many decisions about the testing program.

a. Language usage. This category includes terms (e.g., words, phrases, names) that might, through either their denotations or connotations,

be considered offensive to persons who are of a certain race, ethnicity, gender, age, physical or mental condition, or geographical, economic, or cultural background.

b. Stereotyping. This category refers to the practice of automatically ascribing certain characteristics (e.g., physical, mental, personal, occupational, avocational, cultural, economic) to every member of a group.

c. Representational fairness. This category deals with the underrepresentation or overrepresentation of certain groups in portrayals of activities, roles, relationships, and settings.

d. Content inclusiveness. This category pertains to the potential omission of certain relevant but nontraditional topics, facts, interpretations, and areas of information from the knowledge base embodied in a test (e.g., the contributions of women and Blacks to U.S. history and culture).

3. Objective reviews. After they were drafted by NES, the frameworks and objectives underwent a series of reviews in Illinois. They were first reviewed by ISBE curriculum specialists and then revised to reflect suggestions received from this group. Next they were submitted to the 51 Content Advisory Committees at the Content and Bias Review Conference, a major series of meetings in Illinois. Each Content Advisory Committee had the opportunity to review and revise the draft frameworks.

The committee members--including the ISBE Equity Resource Persons who were serving with the members--were instructed to review the objectives according to several criteria, one of which was freedom from bias.

Following the reviews by the Content Advisory Committees, the frameworks were submitted to the Bias Review Committee. This committee had an opportunity to review the objectives as they had been revised by the Content Advisory Committees and to hear from the Equity Resource Persons a summary of the discussions and decisions relating to the issue of bias that had transpired during the Content Advisory Committee meetings. The Bias Review Committee recommended changes in particular objectives and also identified a number of test fields and objectives that would be of particular interest during test development for reasons of their sensitivity to potential bias or for the opportunity they afforded to address content that would reflect issues of cultural diversity on the tests.

After all reviews, the ISBE considered the objectives, revisions, suggestions, and recommendations of all groups and made determinations about the content of the objectives in each field.

4. Content Inventory. As a further verification that the objectives upon which the testing program would be built reflected actual educational practice in Illinois, the objective sets as approved by the ISBE were reformatted into Content Inventory surveys and administered between

November 1986 and January 1987 to over 8,500 classroom educators and 3,800 teacher educators across all fields. The educators were asked to rate, on a 1-to-5 Likert-type scale, the importance of each objective to the job of an entry-level educator in their fields.

The ISBE made efforts to ensure the participation of minorities in the Content Inventory survey. Oversampling of minority groups was employed as a strategy to ensure adequate representation. The strategy was successful. The judgments of minority educators were therefore a part of the content verification process.

In reviewing the ratings of objectives by the participants in the Content Inventory, the ISBE considered the issue of emerging content. The ISBE decided in a few cases to retain in the frameworks an objective that had received relatively low ratings on the Content Inventory. Objectives of this sort represented new or emerging areas of content, including changes in subject-matter content or emphasis, issues of safety, and subject matter or topics related to cultural diversity. The ISBE considered it important to retain these objectives in the content frameworks.

5. Test item development. Once the objectives for the pilot test had been selected by the ISBE, item development began. Item development teams (consisting of item writers/editors and content and bias reviewers) were assembled by NES and given both an orientation to the project and training in item development. The orientation and training sessions

stressed the issue of bias and included reference to principles and emphases derived from the Illinois reviews of the objective sets, particularly the review by the Bias Review Committee.

The manual Bias Issues in Test Development was used in the training sessions and was distributed to the members of the item development teams with instructions to read it and use it for reference on the important issue of bias prevention.

6. Pre-pilot test reviews. The approximately 8,000 items for the ICTS were scheduled for pilot testing in April 1987; before the pilot test, a group of 574 items was identified for bias reviews. These items corresponded to objectives that the Bias Review Committee had chosen as being of particular interest either for their potential for biased treatment or for their ability to address nontraditional, culturally diverse content.

These items were reviewed first by an equity advisory group convened by National Evaluation Systems, composed of educators from the Amherst, Massachusetts, area skilled in the detection and prevention of bias in educational contexts. As a second step, the items, revised to reflect the comments of the local equity reviewers, were submitted to a two-stage review process in Illinois by the ISBE. The first stage of the review was performed by ISBE Equity Resource Persons; during a two-day session, these reviewers noted potential issues of bias and recommended changes to be considered during the second stage of the review. The

second stage was performed by the Bias Review Committee in its own two-day session. The members of the Bias Review Committee who performed the review considered the suggestions and comments of the ISBE Equity Resource Persons, made final recommendations to the ISBE on those comments and suggestions, and added some further comments on the items. Each recommendation of the Bias Review Committee took one of three forms: (1) a proposed revision; (2) a request for a rewrite; or (3) a request for verification of content by either NES or the appropriate Content Advisory Committee.

7. Pilot testing. A pilot test of all draft test items for 53 of the 54 ICTS test fields (one test, Russian, was on a later development schedule) was conducted in April 1987. More than 2,400 individual pilot tests were administered in 20 administration sessions at 14 sites across Illinois. The tests were administered to volunteers from the student populations of Illinois teacher preparation institutions. Efforts that the ISBE undertook to ensure the participation of minority students included seeking the involvement of minority student advisors on all the campuses from which students were to be drawn. The ISBE Bias Review Committee was also helpful in publicizing the pilot test and the importance of minority participation. Organizations such as the Illinois Committee on Black Concerns in Higher Education, the Illinois Association of Educational Opportunity Program Personnel, and the Latinos Advancing Education were contacted by the ISBE and asked to lend their support and encouragement to the pilot test. The results were positive: minority participation in the pilot test (about 11%)

matched the representation of minority group members in the student populations of the teacher education programs covered by the ICTS (about 10%).

The ISBE Technical Review Committee had reviewed in advance the plans for the pilot test, which included reporting to the Bias Review Committee and the Content Advisory Committees group performance (i.e., percent correct) comparisons for Blacks and Whites and males and females. After the pilot test was completed and participation rates were examined, the Technical Review Committee advised that group performance comparisons be reported to the Content Advisory Committees only in cases where there were 25 or more participants in each comparison group. Due to the nature of the certification areas, the samples of students who took the examinations in many fields were small. This is because the populations of students preparing for certificates in many areas were small also (e.g., Italian and other foreign languages, physics and other sciences). For this reason, Black and White group comparisons were reported to the Bias Review Committee and the Content Advisory Committees in only two of the 53 fields: Basic Skills and General Administrative.

8. Item review meetings. Once the pilot test had been completed, the items were prepared for item review activities in Illinois. These reviews involved ISBE staff members, the Bias Review Committee, and the Content Advisory Committees (which generally included Equity Resource Persons among them). Copies of the Bias Issues in Test Development

manual were sent to the members of the Content Advisory Committees before the review meetings.

At the meetings, reviewers were oriented to the task, including the use of bias as a major item review criterion. All reviewers were given copies of the items as they had appeared during the pilot test and a summary data report. This report contained, for each test, demographic and other background statistics describing the sample of examinees and, for each item, the number of examinees responding, a percent correct figure, point biserial correlations (item to objective and item to total test), and the response distribution (expressed as numbers if the total number of examinees was less than 25, and as percents otherwise). For the two test fields where there were 25 or more examinees in each reporting group (Basic Skills and General Administrative), Black-White and Male-Female comparisons were also reported, displaying the p-values and response distributions for each group. In addition, items in these two fields were flagged if they exhibited significant differences in the two groups' performance according to the Angoff transformed item difficulty method of statistical bias detection, modified using residualized deltas (Shepard et al., 1985).

The first group to perform its item review was the Bias Review Committee in a two-day meeting immediately preceding the Content Advisory Committees' meetings in July 1987. The Bias Review Committee first considered the items and data in the two fields for which there were comparative statistics for Black and White examinees, Basic Skills and

General Administrative. Recommendations for changes or queries for the Content Advisory Committees were recorded by NES facilitators working with the Bias Review Committee.

After the review of the Basic Skills and General Administrative fields, the Bias Review Committee examined other items, especially those corresponding to the objectives that the committee had identified as being of special interest during the objective review phase of the program. Some of these items had been revised before the pilot test in accordance with the committee's recommendations. All were given a second review in the light of the pilot test statistics (i.e., those for the entire sample) and some further suggestions and recommendations were made, which were to be considered by the Content Advisory Committees when they convened.

Following the Bias Review Committee meeting, the Content Advisory Committees met for two days each to review the draft test items and pilot test statistics for their fields. During their orientation sessions, they were reminded of the importance of bias prevention in the ICTS, informed of the work that the Bias Review Committee had completed previously, and presented with the review criteria they would be expected to apply during their consideration of the draft test items and the pilot test statistics. One of these criteria was freedom from bias.

The first task of the Content Advisory Committees was to select objectives from the frameworks that would actually be tested in the initial test administrations of the ICTS. The committees were given data on the rankings of objectives according to: (1) judgments made by the Content Advisory Committee members individually during the objective review conference the previous September; (2) the ratings of the school personnel who responded to the Content Inventory surveys; and (3) the ratings of the teacher education faculty who responded to the Content Inventory surveys. The committees were instructed to consider the data (especially their own prior judgments and the ratings of the school personnel) and the ISBE selection of objectives reflecting emerging areas of content.

After the selection of objectives for testing was accomplished, the Content Advisory Committees reviewed the individual test items (with pilot test data) against several criteria, including freedom from bias. The comments, queries, recommendations, and suggestions of the Bias Review Committee pertaining to each test field were brought to the attention of the Content Advisory Committee for that field by the NES facilitator working with the committee. These Bias Review Committee contributions were given consideration by the members of the committees, including the ISBE Equity Resource Persons serving as committee members.

The ISBE reviewed the work of the Content Advisory Committees, particularly as it related to bias prevention. Specifically, the ISBE reviewed the responses of the Content Advisory Committees to the bias-related

recommendations of the Bias Review Committee. In some cases, the ISBE decided to make changes in the items to reflect more closely the recommendations of the Bias Review Committee while still ensuring item accuracy. After the ISBE review, nearly all of the bias-related suggestions of the Bias Review Committee were adopted.

The result of the reviews was a set of revised test item banks for use in the actual test administrations of the ICTS. These items had undergone the scrutiny of several levels of reviewers with differing perspectives and had benefited from each stage of review.

DISCUSSION

Statistical and judgmental reviews for bias. In the ICTS, both statistical and judgmental procedures were used to detect bias; however, due to the nature of teacher certification testing and the small pools of examinees available in many fields, statistical procedures to detect possible Black-White bias in the draft test items could be applied to only two pilot test fields. Judgmental procedures, which would have been used in any case, thereby attained wider application. In many fields, judgments of educators were the only means available for detecting potential bias in particular items, and this is certain to remain the case for actual ICTS test administrations in many fields.

From a theoretical as well as a practical viewpoint, judgmental procedures are appropriate. Even in situations where the numbers of examinees permit some reliance on statistics, the role of statistics at the item level is generally to focus the attention of reviewers to a particular item so that they can exercise their judgment on it. If the statistically targeted item appears truly biased, the experts can revise or remove it; if in their judgment it is unbiased, they may elect to retain it, perhaps making it the target of future reviews.

Multiple judgmental reviews. Test objectives and items drafted for the ICTS were reviewed at many stages by persons of diverse backgrounds belonging to several groups with varying perspectives and roles. The resulting diversity of reviews in our view benefited the program.

One implication of a multiplicity of reviews is of course disagreement. An effective means of adjudicating contradictory opinions must be applied in such cases. For the ICTS, the ISBE was clearly the adjudicating body. In making its decisions, the ISBE was guided by its legislative mandate to develop a bias-free testing program. In most instances where a judgment call had to be made, the ISBE assigned considerable weight to the opinions of the bias specialists on the Bias Review Committee or the ISBE equity staff.

Sequential bias reviews. The many reviews for bias occurred not simultaneously but sequentially. This was significant and beneficial both because the sequential reviews served as multiple screens through which

the materials were made to pass and because the observations, recommendations, and principles of one group of reviewers could be passed on to assist subsequent reviewers.

An expanded definition of bias. The four-part definition of bias that was a feature of the Bias Issues in Test Development manual influenced decisions during the course of the development of the program. In particular, the concept of content inclusiveness caused test developers to expand their vision of important content necessary for educators in Illinois.

Curricular change. The process of test definition and development undertaken for the ICTS is expected to generate some revisions to curriculum in Illinois teacher preparation programs. In each certificate field, the important content that Illinois educators should know has been clarified through advisory committee work and the Content Inventory survey; this content is now to be published as a set of test objectives in each field. It is likely that preparation programs will respond to the inclusion of new and emerging areas of content, such as topics related to cultural diversity, by revising their curricula to address these areas. Thus, the definition of test content, which is based on required job content, is expected to provide valuable feedback to the preparation programs.

Item writing principles for bias prevention. In developing items for the tests of the ICTS, the consistent focus on bias generated a collection of item writing principles. In general, they are in keeping with traditional

item writing principles; however, some are extensions of these principles. A few of these are summarized and discussed below.

1. Tests and test items can communicate to examinees assumptions that may appear to be held by the test sponsors. In developing tests, care should be exercised to ensure that the items are as "assumption free" as possible. For example, a history objective on developing nations can be represented by test items that focus exclusively on the nations' perceived problems and difficulties; such a negative emphasis can appear to demean or patronize these nations and any persons who trace their heritage to them. By conscious attention to item content, wording, and emphases, the same countries can be presented more positively while still appropriately addressing the same domain of content.

   Through a similar lack of attention, students with special needs can be depicted as sources of repeated classroom difficulties, drains on a teacher's energies, and potential disturbances to their peers. Students of different cultures can be portrayed as sources of potential conflict. History can be presented as a succession of traditional events exclusively involving members of the majority culture as agents and members of minority cultures (when they are mentioned at all) as passive recipients of actions.

   This type of bias is subtle because it is based on views that are virtually unconscious to many people and often presented subliminally.

Such views must be raised to the level of consciousness if they are to be avoided in test items.

2. Item development should involve persons from a variety of backgrounds. Few persons can serve as "universal bias detectors." Because some forms of bias may be very subtle and ingrained, persons of different backgrounds (i.e., persons with differing ingrained views and values) should be involved in item development.

3. In some cases, distractors should <u>not</u> be common misconceptions. The generally sound notion of using as incorrect responses the common misconceptions of ill-informed examinees may in some cases hinder bias prevention. For example, in an item about the effects of mainstreaming a student with a learning disability in a classroom, some common misconceptions about learning disabilities could, if permitted onto the test even as distractors, be offensive not only to students with learning disabilities but also to other students.

4. In cases where "common misconception" distractors are considered acceptable, another caution to apply is to avoid using them to set traps for a single segment of the examinee population (e.g., in a group of items testing standard English grammar, using distractors that predominantly mirror the oral patterns of only one particular group in the population). In other words, if items are to use "common misconceptions" as distractors, it would be wise to ensure that the

misconceptions are indeed distributed generally among the examinee population and not disproportionately among a segment of that population.

5. Items whose stems or stimuli present anecdotal situations should be reviewed carefully to ensure that they present no issues of stereotyping or of representational fairness, particularly with regard to gender roles. It is all too easy for a series of items spread across a full-length test to present a stereotypical image of, for example, male students in active, aggressive, or troublemaking roles and female students in passive, receptive, or cooperative roles. Often the effect is cumulative and therefore subtle.

CONCLUSION

In this paper, we have described the Illinois Certification Testing System and the methods employed throughout its development to prevent the introduction of bias into the program. The methods involve a combination of statistical procedures and judgmental reviews, with the latter predominating somewhat during the development stage because of the numbers of examinees in the certification test fields. As the program moves into full implementation, the use of additional statistical methods of bias detection will be considered, including item response theory and the

Mantel-Haenszel technique. However, whatever statistical procedures are selected, the use of judgmental reviews is sure to continue.

We have also presented a series of qualitative findings from our experience with the Illinois program. We hope our recapitulation of them can be helpful to others who are engaged in similar programs in which bias prevention is of great importance.

APPENDIX
ILLINOIS CERTIFICATION TESTING SYSTEM
TEST FIELDS

Code	Field	Grade Level
802	Early Childhood	pre K-3
803	Elementary	K-9
804	Educable Mentally Handicapped	K-12
805	Trainable Mentally Handicapped	K-12
806	Learning Disabilities	K-12
807	Social/Emotional Disorders	K-12
808	Deaf and Hard of Hearing	K-12
809	Speech and Language Impaired	K-12
810	Blind and Partially Sighted	K-12
811	Physically Handicapped	K-12
812	Media	K-12
813	Reading	K-12
814	English as a Second Language	K-12
815	Guidance	K-12
816	School Psychology	K-12
817	School Social Work	K-12
818	School Nurse	K-12
819	General Supervisory	K-12
820	General Administrative	K-12
821	Superintendent	K-12
822	Chief School Business Official	K-12
823	History	6-12
824	Social Science	6-12
825	English	6-12
826	Spanish	6-12
827	French	6-12
828	German	6-12
829	Hebrew	6-12
830	Italian	6-12
831	Latin	6-12
832	Russian	6-12
833	Dance	6-12
834	Speech	6-12
835	Biological Science	6-12
836	Mathematics	6-12
837	Chemistry	6-12
838	Computer Science	6-12
839	General Science	6-12
840	Physical Science	6-12
841	Physics	6-12
842	Health	6-12
843	Health Occupations	6-12
844	Home Economics	6-12
845	Industrial Arts	6-12
846	Agriculture	6-12
847	Business/Marketing/Management	6-12
848	Art	K-12
849	Music	K-12
850	Physical Education	K-12
851	Art	6-12
852	Music	6-12
853	Physical Education	6-12
854	Theatre Arts	6-12
896	Basic Skills	N/A

References

Allan, R. G., Nassif, P. M., & Elliot, S. M. (Eds.). (1987). Bias issues in teacher certification testing. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

AERA/APA/NCME Committee. (1985). Standards for educational and psychological testing. Washington, D.C.: American Psychological Association, 1985.

Angoff, W. H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias (pp. 96-116). Baltimore, MD: Johns Hopkins University Press.

Berk, R. A. (Ed.). (1982). Handbook of methods for detecting test bias. Baltimore: Johns Hopkins University Press.

Burrill, L. E. (1982). Comparative studies of item bias methods. In R. A. Berk (Ed.), Handbook of methods for detecting test bias (pp. 161-179). Baltimore, MD: Johns Hopkins University Press.

Holland, P. W., & Thayer, D. T. (1986). Differential item functioning and the Mantel-Haenszel procedure. Princeton, NJ: Educational Testing Service, Research Statistics Group.

Ironson, G. H. (1982). Use of chi-square and latent trait approaches for detecting item bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias (pp. 117-160). Baltimore, MD: Johns Hopkins University Press.

Shepard, L. A. (1982). Definitions of bias. In R. A. Berk (Ed.), Handbook of methods for detecting test bias (pp. 9-30). Baltimore, MD: Johns Hopkins University Press.

Shepard, L. A., Camilli, G., & Williams, D. M. (1985). Validity of approximation techniques for detecting item bias. Journal of Educational Measurement, 22(2), 77-105.

Tittle, C. K. (1982). Use of judgmental methods in item bias studies. In R. A. Berk (Ed.), Handbook of methods for detecting test bias (pp. 31-63). Baltimore, MD: Johns Hopkins University Press.